

TD9 : Hashage

1 Hashage avec adressage ouvert

Les tables de hashage utilisent généralement deux variantes :

l'adressage fermé : les collisions dans la table sont gérées en stockant dans chaque case de la table une liste chaînée de toutes les valeurs hashées sur cette case ;

l'adressage ouvert : les collisions sont gérées en poursuivant le hashage en fonction du nombre d'essais réalisés jusqu'à trouver une case libre pour stocker la valeur insérée.

Pour l'adressage ouvert, différentes stratégies de recherche de case libres sont utilisées. Pour la suite, nous noterons x la valeur insérée, et $h(x)$ la valeur de hashage associée. Les plus connues sont :

le rehashage linéaire : lorsqu'une case est occupée, les cases suivantes sont étudiées une à une jusqu'à en trouver une libre. Ainsi, après i collisions, la case examinée est la case

$$l(x, i) = h(x) + i$$

le rehashage quadratique : lorsqu'une case est occupée, les cases suivantes sont étudiées, en augmentant petit à petit la longueur du pas. Après i collisions, la prochaine case sera i cases plus loin. Ainsi, après i collisions, la case examinée est la case

$$q(x, i) = h(x) + i^2$$

le double hashage : pour éviter que les valeurs ne se concentrent dans la même zone du tableau, le double hashage utilise une seconde fonction de hashage g (éventuellement la même que la première) pour déterminer avec quel pas avancer dans la table pour chercher de nouvelles cases libres. La case examinée après i collisions est donc :

$$d(x, i) = h(x) + ig(x)$$

Bien sûr, les indices de cases sont à considérer modulo la taille de la table.

Considérons une table de hashage de taille 13, et les clés suivantes à insérer :

- Athènes, $h(\text{Athènes}) = 1$
- Lisbonne, $h(\text{Lisbonne}) = 4$
- Londres, $h(\text{Londres}) = 4$
- Madrid, $h(\text{Madrid}) = 7$
- Amsterdam, $h(\text{Amsterdam}) = 10$
- Berlin, $h(\text{Berlin}) = 8$
- Paris, $h(\text{Paris}) = 7$
- Rome, $h(\text{Rome}) = 10$

- a) Insérez ces valeurs dans la table en utilisant le rehashage linéaire.
- b) De même en utilisant le rehashage quadratique.
- c) De même en utilisant le double hashage, et avec $g = h$ comme seconde fonction de hashage.

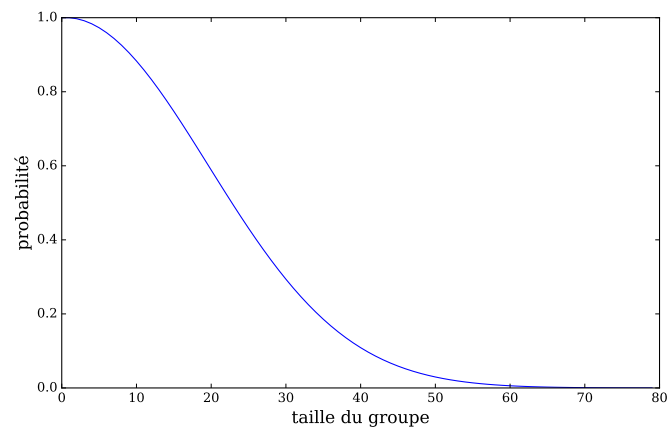
2 Analyse de la probabilité de collision

Dans cette partie, nous allons analyser la probabilité d'obtenir une collision sur n évaluations d'une fonction de hashage uniformément distribuée pour une table de taille m . Illustrons ce calcul sur un groupe de personnes, en utilisant leurs dates d'anniversaire comme valeur de hashage. Nous supposons donc que les dates de naissances sont uniformément réparties sur l'année, et ne présentent pas de variations saisonnières.

- d) En considérant qu'une année fait 365 jours, et que les dates de naissances sont réparties uniformément sur toute l'année, quelle est la probabilité pour une personne de naître un 17 octobre ?

- e) Considérons maintenant que nous avons un groupe de deux personnes. Quelle est la probabilité pour la seconde personne d'avoir le même anniversaire que la première ? La probabilité d'être née un jour différent ?
- f) En supposant que les deux personnes précédentes sont nées des jours différents, et qu'une troisième personne rejoint le groupe. Quelle est la probabilité pour cette personne d'avoir un anniversaire différent des deux précédentes ?
- g) Pouvez vous généraliser pour un groupe de n personnes rejoint par une personne supplémentaire ?
- h) Pour que tout un groupe de n personnes ait des dates d'anniversaire différentes, en numérotant les personnes, il faut que la seconde personne n'ait pas le même anniversaire que la première **et** que la troisième n'ait pas le même anniversaire que les deux précédentes **et** ainsi de suite. Quelle est la probabilité pour que toutes ces conditions soient remplies à la fois ?

Voici la courbe correspondant à la probabilité que tous les membres du groupe aient des anniversaires distincts en fonction de la taille du groupe.



- i) La promotion du cours est de 105 personnes. Parieriez vous plutôt sur le fait qu'elles aient des anniversaires distincts, ou qu'au moins deux personnes aient la même ?
- j) Et vous concernant ? Quelle est la probabilité qu'un autre étudiant de la promotion ait la même date d'anniversaire que vous ?